

Evolutionary role of genomic DNA physical properties in the realization of its elements' biological functions

Mukhina K.¹, Osypov A.^{1,2*}

¹ *Institute of Theoretical and Experimental Biophysics of RAS, Pushchino, Russia*

² *Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow, Russia*

* *aosypov@gmail.com*

Key words: DNA electrostatics, transcription regulation, genome evolution

Motivation and Aim: DNA is highly charged and thus electrostatic interactions with proteins are important in the regulation of transcription. The electrostatic potential is distributed unevenly along the DNA, correlates with GC %, the sequence and its context [1]. Transcription factor binding sites are located in regions with a high electrostatic potential, in areas many times the size of the protein. The distribution of the electrostatic potential on the surfaces of transcription factor molecules reflects the structure of binding sites. Promoters on average have a significantly increased value of the electrostatic potential, the distribution of which correlates with their biological properties. Rho-independent terminators show the same increase in EP, apparently serving to slow down the polymerase to form the terminator hairpin on RNA. Thus, the distribution of the electrostatic potential is an important factor of natural selection that affects the evolution of DNA regulatory regions [1, 2].

Hundreds of transcription factor binding sites lie in protein coding regions, their percentage is not very high, but significant for cell functioning. Some protein coding regions also contain promoters for genes mostly located on the opposite strand.

Methods and Algorithms: DEPPDB and its tools [1, 2] were used to carry out the analysis.

Results: Bacteria belonging to such different evolutionary lineages as Proteobacteria (*E. coli*, representatives of the genera *Salmonella*, *Shigella*, *Yersinia*, *Shewanella*), Firmicutes (*B. subtilis*), and Actinobacteria (*Corynebacterium glutamicum*) were found to have a bias relative to the average distribution of codons and even amino acids around promoters and transcription factor binding sites spanning several hundred codons. This shift is due to the physical properties of the codons in question, providing the correct electrostatic attractors for proteins that regulate transcription (transcription factors and RNA polymerase). The deviation of occurrence from the average level reaches 300 % for codons and 200 % for amino acids. The picture is especially pronounced for such amino acids as glycine, serine, lysine and aspartic acid.

Synonymous codons have, as a rule, the opposite tendency to change in frequency depending on the calculated value of the electrostatic potential they form, the more pronounced, the higher the potential deviation and the frequency of occurrence of the codon and the amino acid itself (leucine is a striking example).

Conclusion: The data obtained require a serious revision of the concept of molecular evolution. The revealed non-synonymity of synonymous substitutions can lead to incorrect assessment of the evolutionary fate of sequences, including incorrect use of the molecular clock and erroneous estimates of the specificity of mutations. The shift in amino acid preferences leads to an even more important revision of the concept of the

proteins natural selection. It substantiates the idea of DNA not only as a text for the implementation of the first step of the Central Dogma of Molecular Biology, but as a complex organ of heredity that fulfills different, sometimes contradictory, requirements, such as packaging and storage, regulation of the implementation of hereditary information, and, finally, encoding proteins and functional non-coding RNAs, whose properties, in turn, depend on the primary structure. These ideas correspond to the ideas of the great evolutionist I.I. Schmalhausen about non-redundant duplication, which he developed in relation to the structures and functions of the animal organism.

Apparently, DNA electrostatics plays an important and universal role in the regulation of prokaryotic transcription, affecting the probability of protein binding and their positioning. The effect of EP on the functioning of the genome is closely related to the effects of other physical and structural properties, first of all, thermal stability, bending, and supercoiling. They can interact both at the stage of their formation and in the process of functional action. Since the physical properties of DNA often depend on the sequence on a statistically averaged basis, information is lost at the stage of their formation. This can work in coding regions due to the redundancy of the genetic code, and different physical properties can operate without interfering with each other.

All this can affect horizontal gene transfer, the evolution of transcription regulation systems, and contribute to AT enrichment of regulatory regions of the genome, and also allows us to put forward a hypothesis about the evolutionary originative role of DNA physical properties in the formation of the transcription regulation system. The regularities of the formation of EP and other physical properties affect such fundamental problems as the 2nd Chargaff rule, the redundancy of the genetic code, the evolutionary non-neutrality of synonymous substitutions and serve as a justification for the fundamental idea of the DNA phenotype, defining the principles of biophysical bioinformatics.

Thus, the electrostatic properties of genomic DNA in combination with other physical properties are very important for the evolution of the realization of its biological functions, primarily for the regulation of transcription. The task of presenting the relevant information is designed to be solved by the database DEPPDB developed by us, available for academic use at <http://deppdb.iteb.ru>.

Acknowledgements: The study is supported by the RFBR (16-04-01865).

References

1. Osypov A.A. et al. DEPPDB – DNA electrostatic potential properties database. electrostatic properties of genome DNA. *JBCB*. 2010;8(3):413-425.
2. Osypov A.A. et al. DEPPDB – DNA Electrostatic Potential Properties Database. Electrostatic Properties of Genome DNA elements. *JBCB*. 2012;10(2):1241004.